日本語形態素解析システムのための形態素文法

渕　武志[†,††] 米澤　明憲[†]

本稿では，動詞の語尾変化について体系的な扱いが可能な派生文法に基づいて，日本語形態素解析システムのための形態素文法を記述した。但し，派生文法における音韻単位での扱いを日本語の文字単位の扱いに変更する方法を示し，より形態素解析処理に適した形で記述した．さらに，これを実働する形態素解析システムに適用し，EDRコーパスと比較することによって精度を測定した．

キーワード: 形態素解析, 日本語, 形態素, 派生文法, 自然言語処理

A Morpheme Grammar for Japanese Morphological Analyzers

TAKESHI FUCHI[†,††] and AKINORI YONEZAWA[†]

This paper shows a morpheme grammar for Japanese morphological analyzers. The grammar bases on Derivational Grammar in which the inflection of Japanese verbs can be managed systematically. Though Derivational Grammar uses phonological description, this paper shows a method to process Japanese character strings directly. A Japanese morphological analyzer with our grammar is evaluated with EDR corpus.

KeyWords: *Morphological Analysis, Japanese, Morpheme, Derivational Grammar, Natural Language Processing*

1　まえがき

形態素解析処理は自然言語処理の基本技術の一つであり，日本語の形態素解析システムも数多く報告されている (吉村, 日高, 吉田 1983)(久光・新田 1990)(中村, 吉田, 今永 1991)(宮崎・高橋 1992)(木谷 1992)(久光・新田 1994a)(丸山・荻野 1994)(松本, 黒橋, 宇津呂, 妙木, 長尾 1994)(Nagata 1994)．しかし，使用している形態素文法について詳しく説明している文献は少ない．文献 (宮崎・高橋 1992) では三浦文法 (三浦 1975) に基づいた日本語形態素処理用文法を提案しているが，品詞の体系化と品詞間の接続ルールの記述形式の提案のみに留まり，具体的な文法記述や実際の解析システムへの適用にまでは至っていない．公開されている形態素解析システム JUMAN(松本他 1994) では，形態素文法は文献 (益岡・田窪 1992) に基づくものであった．その他の文献は解析のアルゴリズムや，固有名詞や未知語の特定機能に関する報告で，使用された形態素文法については述べられていない．

† 東京大学理学部 情報科学科, Department of Information Science, Faculty of Science, University of Tokyo
†† 現在，ＮＴＴ情報通信研究所, NTT Information and Communication Systems Laboratories

言語学の分野で提案されている文法を形態素解析に適用する場合の問題点は，品詞分類が細か過ぎる点と，ほとんどの場合，動詞の語尾の変化について全ての体系が与えられていない点である．言語学の分野では文の過剰な受理を避けるように文法を構築することによって，日本語の詳細な文法体系を解明しようとするので，品詞分類が細かくなるのは当然である．しかし，そのために，文法規則も非常に細かくなり，形態素の統一的な扱いも難しくなる．そこで，本文法では，「形態素解析上差し支えない」ことを品詞の選定基準とする．つまり，ある品詞を設定しないが為に，ある文節に関して構文上の性質に曖昧性が生じる場合に，その品詞を設定する．そして，過剰な受理を許容することと引き替えに，できる限り形態素を統一的に扱う．

従来の多くの文法では活用という考え方で動詞の語尾変化を説明するが，それらの活用形についての規則は，個々の接尾辞について接尾可能な活用形を列挙するという形になっている．例えばいわゆる学校文法では，「書か」はカ行五段活用動詞「書く」の未然形であり，否定の接尾辞の「ない」や使役の接尾辞の「せる」が接尾する等の規則が与えられる．さらに一段活用動詞には「せる」ではなく「させる」が接尾する等の規則があり，規則が複雑になっている．そのため，それらの複雑な規則を吸収するために活用形を拡張し，「書こう」を意志形としたり，「書いた」を完了形とするような工夫がなされる．しかし，このように場当たり的に活用形を拡張すると活用形の種類が非常に多くなり，整合性を保つための労力が大きくなる．

日本語形態素処理における動詞の活用の処理については文献 (Hisamitu and Nitta 1994; 久光・新田 1994b) に詳しい．そこでは，音韻論的手法 (Bloch 1946; 寺村 1984)，活用形展開方式，活用語尾分離方式が紹介され，新たに活用語尾展開方式[1]が提案されている．音韻論的手法は，子音動詞の語幹と屈折接辞を音韻単位に分解し，屈折接辞の音韻変化の規則を用いて，活用を単なる動詞語幹と屈折接辞の接続として捕らえる．しかし，これまでの音韻論的手法では，子音動詞についての知見しか得られていなかったために，子音動詞に接尾する接尾辞と母音動詞に接尾する接尾辞を別々に扱わなければならなかった．また，音韻単位で処理する必要があると考えられているため，文献 (Hisamitu and Nitta 1994; 久光・新田 1994b) でも，処理の効率が落ちるとされている．一方，活用形展開方式，活用語尾分離方式，活用語尾展開方式は何れも伝統的な学校文法に基づいている．活用形展開方式は，各動詞についてその活用形を全て展開して辞書に登録し，それぞれ接尾辞との接続規則を与えるもので，処理速度の点で有利であるが，登録語数が非常に多くなる上に，接続規則の与え方によっては効率の点でも不利になる可能性がある．活用語尾分離方式は，活用語尾を別の形態素とし，動詞語幹と活用語尾の接続規則および活用語尾と接尾辞の接続規則を与えるもので，動詞の屈折形の解析の際に分割数が多くなり，効率の点で不利である．また，接続規則が非常に複雑になる．活用語尾展開方式は，活用語尾と接尾辞の組み合わせを形態素とし，これらと動詞語幹との接続規則だけを与えるもので，活用語尾分離方式よりも分割数が少なくなり，効率的に有利であるとされてい

1 文献 (Hisamitu and Nitta 1994; 久光・新田 1994b) では，提案方式と呼ばれている．

る．しかし，活用形展開方式，活用語尾分離方式，活用語尾展開方式の共通の問題点は，活用語尾と接尾辞の接続規則が体系的でない点である．特に活用語尾展開方式では，新しい接尾辞を追加する度に10以上ある動詞の活用の型それぞれに対する形態素の展開形を追加しなければならない．また，「られ」「させ」といったいわゆる派生的な接尾辞に対してはさらに多くの展開形を別々の形態素として登録する必要があるはずである[2]．

そこで本文法では，動詞の語尾変化について体系的に扱うことに成功している派生文法 (清瀬 1989) を基にした[3]．派生文法も音韻論的アプローチの文法であるが，従来のものに対して，連続母音と連続子音の縮退，および内的連声[4]という考え方を用いて，母音動詞も子音動詞も同様に扱うことができる．しかし，派生文法は音韻論的手法であるため，形態素解析に適用するには，処理を音韻単位で行う必要があるという問題がある．日本語のテキストを処理するような形態素解析システムでは，文字を子音と母音に分けずに日本語の文字でそのまま処理できた方が都合がよい．本研究では，派生文法における動詞語尾の扱いを日本語の文字単位で処理できるように変更する方法を見い出すことができた．すると図らずも従来の活用という考え方に適合する形になることが判明し[5]，これによって，活用の考えを用いて作られている既存の形態素解析システムに適用することができた．しかも語尾変化についての完全な体系を背後に持つため，新たに認識された語尾変化に対しても活用形を順次増やす必要がなく，対応する形態素を一つだけ辞書に登録すれば済むようになった．事実，「食べれる」といったいわゆる「ら抜き表現」や，「書かす」といった口語的な使役表現などもそれぞれ一つの形態素を追加することで対応できている．このように新しい語尾を簡単に追加できることから，口語的な語尾の形態素を充実させることができ，口語的な文章に対しても高い精度で解析できるようになった．また，「食べさせられますまい」といった複雑な語尾変化も正確に解析できる．

本研究で開発された形態素解析文法は，文字表記された日本語のテキストから言語データを抽出することを主な目的として開発されたものである．従って，日本語の漢字仮名混じりの正しい文[6]を文節に区切り，その文節の係り受けの性質を識別することを最優先した解析用の文法となっている．また，形態素の意味的な面を捨象し，過剰な受理を許容することで，形態素の統一的な扱いをすることに重点を置いている．これはあくまで計算機上へのシステムの構築を容易にするためであり，なんらかの言語学的な主張をする意図はない．さらに過剰な受理を許容する意味で，この文法は解析用の文法といえる．生成等に利用するにはこの過剰な受理が障害になる可能性がある．また，誤りを含む文の識別に用いるのにも問題がある．本形態素文法はあくまで正しい文の解析に特化した文法として位置付ける必要がある．

---

2 この点については文献中には触れられていない．
3 派生文法を基にしたシステムとしては，文献 (西野, 鷲北, 石井 1992) で，何らかの方法で分解した動詞の語尾の構造を派生文法に基づいて解析するシステムについて報告されているが，形態素解析システムへの適用は報告されていない．
4 上記の屈折接辞の音韻変化と同じもの
5 派生文法では日本語における活用の考え方を完全に否定している．
6 一般の日本人が許容できる範囲で正しいという意味で，正式な日本語という意味ではない．

属性名	属性値
主属性	無 | 動 | 形 | 名 | 数 | 時 | 格 | 尾
係属性	無 | 連体 | 連用 | 終止
左隣接属性	無 | 動 | 形 | 名 | 数 | 時 | 接続 | 連体 | 連用
右隣接属性	無 | 動 | 形 | 名 | 数 | 時 | 接続 | 尾

表 1　形態素の属性

本稿では2節で形態素の種類とそれらが満たすべき制約の体系を説明し，3節で動詞の語尾の扱いについて述べる．4節では，それを日本語文字単位の形態素解析向きに変更する方法を示す．さらに，5節では個別の問題がある語尾について述べ，最後にこの形態素文法を形態素解析プログラム JUMAN に適用した場合の解析性能を評価する．なお，われわれが作成した形態素文法の形態素解析プログラム JUMAN への適用事例は，以下の anonymous ftp で入手可能である．但し，評価の際に使用した辞書の一部について配布に制限のあるものは含まれていない．
camille.is.s.u-tokyo.ac.jp /pub/member/fuchi/juman-fuchi

2　形態素の体系

本稿では，形態素文法を品詞間の隣接可能性を示す形で与える．文献 (丸山・荻野 1994) にあるように，形態素文法も正規文法等で記述した方がより細かい記述ができるが，処理効率や形態素解析システムへの適用の関係でこの形に落ちついた．

本文法では形態素に対して表1に示す属性を設定する．逆にこのような属性を付加できるような文字の最小の連鎖を形態素と呼ぶ．また，形態素をその性質によって分類したものを品詞と呼ぶ．但し，システムの解析精度を上げるために，「かどうか」のように幾つかの形態素から合成され，厳密には形態素と言えないものを，一つの形態素として扱う場合がある．形態素の属性は以下のリストで表記する．

[主属性, 係属性, 左隣接属性, 右隣接属性]

主属性は形態素の基本的な性質による分類であり，その形態素を含む文節がどのような「受け」を構成するかを示す．係属性はその形態素で終わる文節がどのような「係り」を構成するかを示す．文節の「係り」の種類としては「連体」と「連用」を設定する．「連体」は主属性が「名」「数」「時」の形態素に対して係り，「連用」は「述語[7]」に対して係る．そして，「連用」の「係り」となる文節の末尾に位置する形態素の係属性の値を「連用」とし，そのような形態素を「連用形」と呼ぶ．同様に「連体」の「係り」となる文節の末尾に位置する形態素の係属性の値を「連体」とし，そのような形態素を「連体形」と呼ぶ．さらに，「係り」を構成しない文節の末尾

7 「述語」は名詞＋名詞接尾辞，動詞語幹＋動詞語尾，形容詞語幹＋形容詞語尾，名詞＋句読点，数詞＋句読点および時詞＋句読点によって構成される．

形態素の属性	品詞名	略記	補足
[名, 無, 無, 無]	名詞	名	人名, 地名, 物名
			動作名 ex. 跳躍
			形状名 ex. 静か
[名, 連用, 無, 無]	連用名詞	名/連用	ex. 道中, 半分, 反面
[名, 連用, 連体, 無]	補助名詞	補名	ex. の, ん, 際
[(名, 時), 連用, 無, 無]	時詞	時	時の名称 ex. 今日, 夏
[(名, 数), 無, 無, 無]	数詞	数	数字 ex. 1 , 一, 壱

表 2　名詞, 数詞, 時詞, 補助名詞

に位置する形態素の係属性の値は「終止」とし, そのような形態素を「終止形」と呼ぶ.

左右の隣接属性は隣接する二つの形態素が満たすべき制約を表している. 文を左から右に記述した場合に, 隣接している形態素の内, 左にある形態素の属性が [X1,Y1,L1,R1] であり, 右にある形態素の属性が [X2,Y2,L2,R2] であったならば, 以下が成り立つ必要がある.

$R1 \in X2 \cup Y2 \cup L2$ かつ $L2 \in X1 \cup Y1 \cup R1$.

大まかには, 左にある形態素の右隣接属性は, その値がすぐ右の形態素の主属性, 係属性, 左隣接属性のいずれかの値と同じである場合に, それらの形態素が隣接可能であることを示す. また, 右にある形態素の左隣接属性は, その値がすぐ左の形態素の主属性, 係属性, 右隣接属性のいずれかの値と同じである場合に, それらの形態素が隣接可能であることを示す. また, 右隣接属性が「接続」である形態素を「接続形」と呼ぶ.

これらの属性は, その取りうる値の組合せの内, 一部は実在しない. 表 2から表 15に実在する品詞を示す. 以下で, それぞれの品詞について説明する. なお, それぞれの品詞名は, なるべく統語的な性質を反映した名前になるように本研究で独自に与えたものである. また, 表中の補足の欄で与える例の内, アルファベットで表記してあるものは, 3節で説明する連続母音縮退, 連則子音の縮退, および内的連声との関連で, そのままでは日本語文字表記にならないものである.

2.1　名詞, 連用名詞, 補助名詞, 時詞, 数詞

表 2に名詞関連の形態素を示す. 名詞の属性は [名, 無, 無, 無] である. 主属性が「名」である事は, 連体の係りを受けることを意味する. また, 隣接属性が左右とも「無」であることは, 名詞が単独で文節を構成可能であることを示す. 係属性が「無」であることは, 名詞自身では係りの種類を指定しないことを示す. 特に, 直後に名詞がくる場合には, 複合して一つの名詞を形成する.

形態素の属性	品詞名	略記	補足
[格, 連用, 名, 無]	格接尾辞/連用形	格尾/連用	格助詞 ex. が, を, に, で
[格, 連体, 名, 無]	格接尾辞/連体形	格尾/連体	属格, 助詞 ex. の, や, か, と

表 3　格接尾辞

名詞はさらに細かく「動作名詞」や「形状名詞」などへの分類が可能である．これらの細分類には「する」が接尾可能であるとか,「な」が接尾可能であるなどの統語的な振る舞いの違いが見られる．しかし，基本的にはこれらの細分類は意味的なものを反映している．従って，発話者がある単語にどのような意味を込めるかによって変動しうるものである．聞き手の立場からは逆にそのような使われ方から発話者が込めた意味を読みとる必要がある．従って，解析の場合には，解析の際に不都合がない限り，このような細分類は必要でないと考える．特に一般的に形容動詞といわれるものも，文献 (時枝 1950) と同様に，形容的な意味合いが強い名詞として，名詞に含める．

連用名詞は属性が［名, 連用, 無, 無］の品詞で，係りの種類を「連用」に指定する名詞の一種である.「半分冗談で言った」の例のように，直後に名詞が来ても複合名詞を形成せず，連用の文節を形成する．但し「冗談半分で言った」のように連用名詞が名詞の後に来る場合には複合名詞を形成する．

補助名詞は述語の連体形の直後のみに現れる[8]という性質以外は連用名詞と同様な振る舞いをする．代表的な補助名詞は「の」で，属性は［名, 連用, 連体, 無］である.「の」と「ん」は，実際には述語の連体形の直後にしか現れ得ないなど，さらに細かい制約があるが，本文法ではそれらの制約を表していない．

時詞は連用名詞の一種であるが,「昨年夏に」などのように時詞が連続した場合には複合すると考え，別に設定した．属性は［(名, 時), 連用, 無, 無］である．主属性が (名, 時) となっているのは，両方の属性を持つことを表す．

数詞は名詞の一種とも見なせるが，数詞のみに接尾する接尾辞が存在し，これを区別しないと曖昧性が生ずる場合がある．そこで，属性を［(名, 数), 無, 無, 無］として名詞とは別に数詞を設定した．

2.2　格接尾辞

表 3に格接尾辞を示す．格接尾辞は名詞に接尾して連用または連体の文節を形成する．格接尾辞が次節の名詞接尾辞と異なる点は，述語を形成しない点である．終止形は述語を形成してしまうため，この品詞には終止形が存在しない．文節に区切る目的からは格接尾辞と名詞接尾

8 従って，左隣接属性が「連体」になる．

形態素の属性	品詞名	略記	補足
[尾, 終止, 名, 無]	名詞接尾辞/終止形	名尾/終止	ex. だ, です, でしょう, だろ
[尾, 連用, 名, 無]	名詞接尾辞/連用形	名尾/連用	ex. で, でして
[尾, 連体, 名, 無]	名詞接尾辞/連体形	名尾/連体	ex. な, だった
[尾, 無, 名, 接続]	名詞接尾辞/接続形	名尾/接続	ex. だ, です, でしょう

表 4　名詞接尾辞

辞を区別する必然性はないが，構文解析での利用を考慮してこのように設定した．しかし「で」などは格接尾辞と名詞接尾辞の両方に所属すると考えられ，しかも形態素レベルで区別する方法がない．また「と」に関しては，連用と連体の両方の用法があると考えられ，これも形態素レベルでは区別ができない．これらについてはその取り扱いを5節で改めて述べる．

2.3　名詞接尾辞

表4に名詞接尾辞を示す．名詞接尾辞は名詞に接尾して述語の文節を形成する．従って「連用」の係りを受ける．

本文法で名詞接尾辞の連体形としている「な」については,「学生なので」と「健康な人」では意味的に異なるものと考えられるが，前者の「の」を補助名詞と考えると，両者とも統語的には同一に扱える．さらに本文法では「な」は話し手が形容的な意味合いを付加したあらゆる名詞に接尾可能であると考える．また,「体が健康な人」の例を考えると「名詞＋な」で述語を形成していることが分かる．従って,「な」は格接尾辞連体形ではなく，名詞接尾辞連体形とする．

接続形の形態素は全て連体形や終止形の形態素と表記が同じであるが，接続形では必ず接続接尾辞が接尾し，逆に連体形や終止形には接続接尾辞が接尾しないので，両者は区別可能である．このことは動詞接尾辞や形容詞接尾辞の接続形についても同じである．

2.4　動詞語幹，動詞接尾辞

表5に動詞関連の形態素を示す．動詞語幹の属性は［動, 無, 無, 尾］であり，右隣接属性が「尾」なので,「尾」の属性を持つものが接尾しなければならない．実際に接尾できるのは動詞接尾辞もしくは派生接尾辞の一部である．動詞語幹は動詞接尾辞を伴って動詞を形成する．動詞語幹には子音で終わるものと母音で終わるものがある．例えば「書く」の語幹は「kak」であり,「食べる」の語幹は「tabe」である．実際に存在する動詞語幹の末尾の子音は，K, G, S, T, N, B, M, R, W の九個である．

動詞接尾辞は動詞語幹または「動」の属性値を持つ派生接尾辞に接尾して，述語を形成する．動詞接尾辞には子音で始まるものと，母音で始まるものがある．例えば「書く」の動詞接尾辞

形態素の属性	品詞名	略記	補足
[動, 無, 無, 尾]	動詞語幹	動	ex. 食べ, 歩 k, 走 r, 思 w
[尾, 終止, 動, 無]	動詞接尾辞/終止形	動尾/終止	ex. ru, ita, you, ina
[尾, 連用, 動, 無]	動詞接尾辞/連用形	動尾/連用	ex. i, ite, eba
[尾, 連体, 動, 無]	動詞接尾辞/連体形	動尾/連体	ex. ru, ita
[尾, 無, 動, 接続]	動詞接尾辞/接続形	動尾/接続	ex. ru, ita

表 5　動詞語幹，動詞接尾辞

形態素の属性	品詞名	略記	補足
[形, 無, 無, 尾]	形容詞語幹	形	ex. 美し, 高
[尾, 終止, 形, 無]	形容詞接尾辞/終止形	形尾/終止	ex. い, かった, かれ
[尾, 連用, 形, 無]	形容詞接尾辞/連用形	形尾/連用	ex. く, くて, ければ
[尾, 連体, 形, 無]	形容詞接尾辞/連体形	形尾/連体	ex. い, かった
[尾, 無, 形, 接続]	形容詞接尾辞/接続形	形尾/接続	ex. い, かった

表 6　形容詞語幹，形容詞接尾辞

は後述するように「ru」であると見なせ,「書きます」の動詞接尾辞は「imasu」であると見なせる．実際に存在する動詞接尾辞の先頭は，A, I, U, E, YO, RU である．さらに動詞語幹に接尾する派生接尾辞にはRA, SA, RE で始まるものがある．

これらの語幹に接尾辞が接尾する場合には，連続母音縮退，連続子音縮退，内的連声という規則的な変換がおこる．その詳細については3節で述べる．

2.5　形容詞語幹，形容詞接尾辞

表 6に形容詞関連の形態素を示す．形容詞語幹の属性は［形, 無, 無, 尾］であり，右隣接属性が「尾」なので,「尾」の属性を持つものが接尾しなければならない．実際に接尾できるのは形容詞接尾辞もしくは派生接尾辞の一部である．形容詞語幹は形容詞接尾辞を伴って形容詞を形成する．

形容詞接尾辞は形容詞語幹または「形」の属性値を持つ派生接尾辞に接尾する．基本的には形容詞接尾辞は動詞接尾辞のような語形の変化はなく，そのままの形で形容詞語幹に接尾する．しかし，形動派生接尾辞「ござ r」が形容詞語幹に接尾する場合にのみ内的連声と呼ばれる語形変化があり，表 7のようになる．例えば,「たか (高)」という形容詞語幹に「ござる」が接続する場合には「taka → takou」と変形され,「たこうござる」となる．

内的連声	具体例
-a → -ou	たこうござる
-i → -yuu	うつくしゅうござる
-u → -ou	さもうござる
-o → -ou	ほそうござる

表 7　形容詞の内的連声

形態素の属性	品詞名	略記	補足
[無, 連体, 無, 無]	連体詞	連体	指示語 ex. その
[名, 連用, 無, 無]	連用詞	連用	副詞 ex. ゆっくり, とても
[無, 連用, 無, 無]	連文詞	連文	接続詞 ex. しかし, ところで
[無, 終止, 無, 無]	終止詞	終止	感動詞, 感嘆詞 ex. おはよう, おや

表 8　連体詞，連用詞，連文詞，終止詞

2.6　連体詞，連用詞，連文詞，終止詞

連体詞は属性が［無, 連体, 無, 無］で，それだけで連体形の文節を構成する形態素である．代表的なものは「その」などの指示を表す形態素である．連体詞はどのような係りも受けない．

連用詞は一般的には副詞と言われるもので，属性が［名, 連用, 無, 無］で，それだけで連用形の文節を構成する形態素である．これには「彼ののんびりにはいらいらさせられる」などに見られる名詞的な用法があるため，主属性を「名」とした．そのため，属性としては連用名詞と同じであるが，名詞の直後に来ても複合名詞を形成しない点が連用名詞とは異なる．例えば「車ゆっくり走らせて下さい」という文では「車ゆっくり」という名詞であるとは受け取られない．その他に,「とてもゆっくり走らせた」の例では「とても」が「ゆっくり」に係ると考えられるが，本文法では「とても」は「走らせた」に係ると考えることにして，連用詞に係る連用詞を設定していない．これに関しては5節でも触れる．

連文詞は一般的には接続詞と言われるもので，文と文をつなぐ働きをする．属性は［無, 連用, 無, 無］で連用詞と似ているが，名詞的な用法がない．また，普通は文頭に現れる．

終止詞は一般的には感動詞や感嘆詞と言われるもので，単独で文を形成し，係り受けを形成しない．

2.7　接頭辞

表 9に接頭辞を示す．接頭辞は，ある形態素に接頭する形態素で，文節全体の係り受けの性質には影響を与えない．本文法では，名詞，時詞，数詞，動詞，形容詞に接頭する接頭辞を設

形態素の属性	品詞名	略記	補足
[無, 無, 無, 名]	名詞接頭辞	頭名	ex. お, ご, 前, 元
[無, 無, 無, 時]	時詞接頭辞	頭時	ex. 翌, 昨, 来
[無, 無, 無, 数]	数詞接頭辞	頭数	ex. 第, 約, 計
[無, 無, 無, 動]	動詞接頭辞	頭動	ex. お, ぶち
[無, 無, 無, 形]	形容詞接頭辞	頭形	ex. お, うすら

表 9　接頭辞

形態素属性	品詞	略記	補足
[名, 無, 名, 無]	名名派生接尾辞	名名	ex. さん, 製, 的
[動, 無, 名, 尾]	名動派生接尾辞	名動	ex. する, できr, ぶr
[形, 無, 名, 尾]	名形派生接尾辞	名形	ex. らし, くさ
[名, 連用, 時, 無]	時名派生接尾辞	時名	ex. 前, 中, 下旬
[名, 無, 数, 無]	数名派生接尾辞	数名	ex. 人, 個, 姉妹
[時, 連用, 数, 無]	数時派生接尾辞	数時	ex. 年, 日, 秒
[数, 無, 数, 無]	数数派生接尾辞	数数	ex. 万, 億, 兆
[(尾, 名), 連用, 動, 無]	動名派生接尾辞	動名	ex. i手, aなさそう
[(尾, 動), 無, 動, 尾]	動動派生接尾辞	動動	ex. sase, rare, imakur
[(尾, 形), 無, 動, 尾]	動形派生接尾辞	動形	ex. aな, iた, iにく
[(尾, 名), 連用, 形, 無]	形名派生接尾辞	形名	ex. そう
[(尾, 動), 無, 形, 尾]	形動派生接尾辞	形動	ex. がr
[(尾, 形), 無, 形, 尾]	形形派生接尾辞	形形	ex. かな
[(尾, 名), 連用, 接続, 無]	接名派生接尾辞	接名	ex. か, かどうか
[(尾, 動), 無, 接続, 尾]	接動派生接尾辞	接動	ex. にすぎ
[(尾, 形), 無, 接続, 尾]	接形派生接尾辞	接形	ex. らし

表 10　派生接尾辞

定する.

2.8　派生接尾辞

表 10に派生接尾辞の一覧を示す. 派生接尾辞は, 名詞や動詞語幹, 形容詞語幹または接続形に接尾して, 品詞を変換し, 新たに語幹を形成する接尾辞である. これらの派生接尾辞は派

形態素属性	品詞	略記	補足
[尾, 終止, 接続, 無]	接続接尾辞終止形	接尾/終止	ex. ぜ, もん, の, か
[尾, 連用, 接続, 無]	接続接尾辞連用形	接尾/連用	ex. し, が, ので, のに
[尾, 連体, 接続, 無]	接続接尾辞連体形	接尾/連体	ex. だろう, でしょう
[尾, 無, 接続, 接続]	接続接尾辞接続形	接尾/接続	ex. だろう, でしょう

表 11　接続接尾辞

生文法 (清瀬 1989) を参考に，本研究で整理，拡充したものである．品詞の名称から個々の派生接尾辞の働きは明らかなので，以下では幾つか注意を要するものについてのみ説明する．

名動派生接尾辞は，名詞に接尾して動詞語幹を形成する接尾辞である．代表的なものが「する」で，これは普通は動作を表す名詞に接尾するが，本文法では発話者が単語に込める意味によって全ての名詞に接尾可能であるとしている．

動名派生接尾辞は，動詞語幹に接尾して名詞を形成する．これは3節で説明する動詞接尾辞の基本接続規則のみに従い，内的連声には従わない．この点は動形派生接尾辞や動動派生接尾辞も同様である．

動動派生接尾辞は，動詞語幹に接尾してまた動詞語幹を形成する．代表的なものは使役の「sase」や受身・尊敬・自発・可能の「rare」である．この動動派生接尾辞は動詞の語幹に次々に接尾して動詞語幹を派生する．例えば「書かせられますまい」という表現は，後述する連続母音縮退や連続子音縮退に注意すれば,「書 k/sase/rare/imas/umai」である．その他に，口語的な表現では「食べさせる」「書かせる」を「食べさす」「書かす」などともいうが，これは「sas」という形態素で説明できる．さらに，可能の意味での「食べれる」「書ける」という表現は「re」という形態素で説明できる．このように最近になって新しく使われるようになったと考えられる表現でも本文法に沿っていることが分かる．

このような派生接尾辞は，個々の形態素の意味に応じて，接尾可能でない場合がある．特に動動派生接尾辞では，その順番に明らかに制約が存在する．しかし，本文法では文法の簡潔性を優先し，それらの制約を反映していない．これらは文生成においては解明すべき重要な問題である．

2.9　接続接尾辞

表 11に接続接尾辞を示す．接続接尾辞は名詞接尾辞や動詞接尾辞，形容詞接尾辞の接続形に接尾して連用形，連体形，終止形，接続形を形成する．これには例えば「かのような」のように幾つかの形態素から成り立っているみなせるものも含まれている．このようなものは，成り立ちは確かに幾つかの形態素の組合せと考えられるが，使用上は一つの接尾辞として振舞う

形態素属性	品詞	略記	補足
[尾, 終止, 終止, 無] [尾, 終止, 連用, 無] [尾, 終止, 名, 無]	末尾接尾辞	尾尾	ex. よ, ね, さ, か

表 12　末尾接尾辞

形態素属性	品詞	略記	補足
[尾, 連用, 終止, 無] [尾, 連用, 連用, 無] [尾, 連用, 名, 無]	引用接尾辞連用形	引用/連用	ex. と
[尾, 連体, 終止, 無] [尾, 連体, 連用, 無] [尾, 連体, 名, 無]	引用接尾辞連体形	引用/連体	ex. との

表 13　引用接尾辞

ので，まとめて一つの形態素として扱う．

名詞接尾辞の「だ」に関連して，この品詞の隣接規則には例外があり，接続接尾辞の連用形と終止形の一部のみが「だ」に接尾可能である．さらに「ので」「のに」に関して個別の例外があり，これは5節で検討する．

2.10　末尾接尾辞

表 12に末尾接尾辞を示す．末尾接尾辞は文の末尾に用いられる接尾辞で，基本的には終止形に接尾し，属性は［尾, 終止, 終止, 無］である．しかし，連用形や名詞に対しても接尾が可能であり，[尾, 終止, 連用, 無]，[尾, 終止, 名, 無］の属性も持つと考えられる．いずれにしても終止形を構成する．末尾接尾辞にさらに末尾接尾辞が接尾することは可能であるが，全ての組合せが見られるわけではない．これは形態素の隣接規則とは別の意味的な制約によるものと考えられるが，解析の立場からは可能な組合せを洗い出す必要はないと考え，全ての末尾接尾辞が隣接可能であるとしている．また，頻出する末尾接尾辞の連続に対しては，一つの形態素として辞書に登録している．

2.11　引用接尾辞

表 13に引用接尾辞を示す．引用接尾辞の基本的な属性は［尾, 連用, 終止, 無］である．これ

形態素属性	品詞	略記	補足
[尾, 連用, 連用, 無] [尾, 連用, 名, 無]	連用接尾辞連用形	用尾/連用	ex. は, も, すら, まで
[尾, 連体, 連用, 無] [尾, 連体, 名, 無]	連用接尾辞連体形	用尾/連体	ex. すらの, までの

表 14　連用接尾辞

形態素属性	品詞	略記	補足
[無, 無, 無, 無]	間投辞	間投	ex. ね, さ, よ
[無, 無, 無, 無]	読点	読点	ex. ,
[無, 無, 無, 無]	句点	句点	ex. . ?!
[無, 無, 無, 無]	括弧	括弧	ex. 「」{}

表 15　間投辞，句読点，括弧

は終止形に接尾して連用形を構成することを意味する．しかし，連用形や名詞に対しても接尾が可能であるため，[尾, 連用, 連用, 無]，[尾, 連用, 名, 無] という属性も持つ．また「首相が辞任したとのニュース」の例のように連体形も存在する．

2.12　連用接尾辞

表 14に連用接尾辞を示す．連用接尾辞連用形の属性は [尾, 連用, 連用, 無] であり，その代表的なものは「は」「も」「すら」「さえ」「だけ」「まで」等である．これらは連用形の形態素に接尾して再び連用形を形成する接尾辞である．これらはさらに名詞にも接尾するので，[尾, 連用, 名, 無] でもある．また隣接規則に例外があり，接続接尾辞連用形には接尾しない．さらに連用接尾辞の隣接可能な組み合わせは全てのものが存在するわけではなく，何らかの意味的な制約によって制限されていると思われるが，本文法では，その制約を追求することはしていない．

連用接尾辞連体形は用いられることは希であるが存在し，属性は [尾, 連用, 連用, 無]，[尾, 連用, 名, 無] である．

2.13　間投辞，句読点，括弧

その他の品詞を表 15に挙げる．その中に，文節の切れ目にのみ置くことができる間投辞がある．これに属する代表的な形態素は「ね」である．また句読点や括弧なども間投詞と形態素としての性質は同じで，これらの品詞の属性は全て [無, 無, 無, 無] である．

2.14　隣接規則の例外

隣接規則の例外は特殊な用法から生まれている．一つは「彼の飛びは最高だった.」のように「動詞語幹＋i」を名詞として扱うものである．これを動名詞と呼ぶ．今一つは，連用形を終止形とみなす用法で，何らかの述語を省略して連用形で文を終わらせてしまう用法である．この場合，本来終止形に接尾するものが連用形に接尾することになる．本研究で使用した形態素解析プログラム JUMAN では隣接規則が自由に記述できるようになっているので，これらの例外的隣接規則も，基本的な隣接規則と同様に記述できた．

3　派生文法による動詞の語尾変化

本節では，派生文法に則して，動詞語幹と動詞接尾辞もしくは派生接尾辞との接続規則を記述する．これを日本語の文字単位で処理するのに適する形に変更する方法については，次節で述べる．

派生文法では，動詞の語幹と接尾辞との接続規則を連続母音の縮退と連続子音の縮退で説明する．連続子音の縮退は従来から指摘されていたものであるが，これに連続母音の縮退という考え方を導入することにより，活用という考え方を用いずとも体系的に現代日本語の動詞の語尾変化を説明できる．具体的には「kak」に「ru」が接尾すると「kak/ru」となるが,「k/r」の部分が連続子音となり，後ろの「r」が縮退し,「kaku」となる．これが連続子音の縮退である．一方「tabe」に「imasu」が接尾すると「tabe/imasu」となるが,「e/i」の部分が連続母音となって後ろの「i」が縮退し,「tabemasu」になる．これが連続母音の縮退である．その他の組み合わせ,「kak」と「imasu」,「tabe」と「ru」の場合はそれぞれ「kak/imasu」,「tabe/ru」となり，子音も母音も連続しないので，縮退せず「kakimasu」,「taberu」になる．以上の接続規則を基本接続規則と呼ぶことにする．

この基本接続規則に加えて，表16に示すような内的連声がある．表中の具体例は完了の接尾辞「ita」との組み合わせで示す．例えば「書く」であれば「kak」に「ita」が接尾するとまず「kak/ita」となり，この「k/it」の部分が内的連声により「it」となるから，最終的には「kaita」となる．この内的連声の唯一の例外が「行く」で,「ik/ita」が「iita」とならずに「itta」となる．注意すべきはこの内的連声は動詞接尾辞が接尾する場合にのみ適用されることで，例えば願望を表す動形派生接尾辞の「i た (い)」では連声しない．

派生文法における動詞の語形変化の扱いの例外が「する」と「くる」の二つの動詞と，これらを使って作られる動詞，さらに「感じる」など「する」が濁音化して変化したと思われる一群の動詞である.「する」「じる」「くる」の語幹変化を表17に示す．これらには同一の接尾辞に対して複数の語幹変化があるものがある．

また,「おっしゃる」「いらっしゃる」「なさる」「下さる」の四つの動詞は，基本的には語幹

音便	具体例
k/it → it	書く kak/ita → kaita 書いた
g/it → id	嗅ぐ kag/ita → kaida 嗅いだ
t/it → tt	待つ mat/ita → matta 待った
n/it → n'd	死ぬ sin/ita → sin'da 死んだ
b/it → n'd	飛ぶ tob/ita → ton'da 飛んだ
m/it → n'd	噛む kam/ita → kan'da 噛んだ
r/it → tt	掘る hor/ita → hotta 掘った
w/it → tt	買う kaw/ita → katta 買った
(k/it → tt)	(例外) 行く ik/ita → itta 行った

(具体例は完了の接尾辞「ita」との組み合わせで示している)

表 16　内的連声

語幹	接尾辞の始まりの部分
s	i-, u-, e-, sa-, ra-
si	anai, yo-
se	a-
sur	u-, e-, re-, ru-

「する」

語幹	接尾辞の始まりの部分
zi	a-, i-, u-, e-, yo-, sa-, ra-, ru-, re-
zur	e-, ru-, re-

「じる」

語幹	接尾辞の始まりの部分
k	i-, u-, e-
ko	a-, yo-, sa-, ra-, ru-
kur	re-

「くる」

表 17　「する」「じる」「くる」の語幹変化

語幹	接尾辞
-r	a-, i-, u-, e-, yo-, sa-, ra-, ru-, re-
-	i, i-

表 18　「おっしゃる」「いらっしゃる」「なさる」「下さる」の語幹変化

が「r」で終わるものと同じであるが，幾つかの語幹の変化がある．それを表18に示す．注意すべきは，内的連声が適用される場合は，内的連声が語幹変化よりも優先されることである．

最後に命令の動詞接尾辞に例外がある．動詞接尾辞「ro」「yo」は母音で終わる動詞語幹にのみ接尾し，「e」は子音で終わる動詞語幹にのみ接尾する．また，語幹が変化する不規則な動詞に関しては命令の形も不規則なものとなる．

動詞の形成における語形の変化は以上の規則で全て説明できるが，このままでは日本語の文字単位で解析するのには向かない．そこでこれらを基礎にして，日本語文字単位の解析向きに変更する方法について次の節で述べる．

4　形態素解析システムへの適用

前節で説明したような形態素を設定すれば，現代日本語の文を構成する形態素を説明できるが，このままでは形態素解析には向かない．なぜならば日本語の文は漢字や平仮名などで記述されているので，特に動詞に関しては一旦文字を子音と母音に分解しなければ解析できないからである．そこで子音や母音に分解せず日本語の文字の単位で解析できるように工夫することを考えた．すると，従来用いられていた活用という考え方に近づき，そのことによって，従来の活用という考え方に沿って作られた形態素解析プログラムに，本稿で示した形態素文法を処理させることができるようになった．

4.1　動詞に関する修正

まず，動詞語幹に接尾する接尾辞の始まりの部分が数種類しかない．さらに動詞語幹の末尾の子音もいくつかの種類に限定されている．そこで，これらの組み合わせを動詞の活用語尾とし，動詞の語幹から末尾の子音を除いた部分を新たに動詞語幹とする．各接尾辞については先頭の部分を隣接型とし，その部分を除いた残りの文字列を新たにその形態素の表記とする．そして，それらの新たな接続規則を設定する．

具体的に，動詞語幹に接尾する接尾辞の始まりは A, I, U, E, YO, T, D, RA, RU, RE, SA の 11 種類である．ここで，T と D は内的連声を形成する it-という接尾辞の始まりを i-と区別したもので，さらに T は清音の内的連声に対応し，D は濁音の内的連声に対応する．そこで個々の動詞語幹に対して，この 11 の活用形を設定することになる．この活用形のパターンは動詞語幹の末尾に応じて決まるが，末尾が母音で終わる場合を A と表記することにすると末尾の種類は A, K, G, S, T, N, B, M, R, W の 10 種類であるから，それに応じた 10 種類の活用型があることになる．さらに例外の活用型を SX, ZX, KX, KKX, RX, IKU で表すことにすると，活用型と活用形の組み合わせは表 19のようになる．

個々の動詞に関しては，語幹の末尾の子音を除いた文字列を「語幹」とし，「食べ」のように語幹が母音で終わる場合には A を，「書 k」のように語幹が子音で終わるものは子音そのものを活用型とする．例えば，「食べ」の場合は「食べ」が語幹で活用型が A，「書 k」の場合は「書」が語幹で活用型が K，「嗅 g」の場合は「嗅」が語幹で活用型が G，「思 w」の場合は語幹が「思」で活用型が W となる．

動詞語幹に接尾するのは接尾辞は動名派生接尾辞，動形派生接尾辞，動動派生接尾辞，そして動詞接尾辞である．これらについて例を挙げると，動名派生接尾辞の「i そう」は隣接型が I 型で表記文字は「そう」となり，動形派生接尾辞の「a な」は隣接型が A 型で表記文字は「な」，動動派生接尾辞の「rare」は隣接型が RA 型，活用型が A 型，表記文字が「れ」となる．接尾辞は例えば「i ます」は隣接型が I 型で，表記文字が「ます」になる．連声するような接尾辞，

活用形 / 活用型	A	I	U	E	YO	T	D	RA	RU	RE	SA
A	*	*	*	*	よ	*	–	ら	る	れ	さ
K	か	き	く	け	こ	い	–	か	く	け	か
G	が	ぎ	ぐ	げ	ご	–	い	が	ぐ	げ	が
S	さ	し	す	せ	そ	し	–	さ	す	せ	さ
T	た	ち	つ	て	と	っ	–	た	つ	て	た
N	な	に	ぬ	ね	の	–	ん	な	ぬ	ね	な
B	ば	び	ぶ	べ	ぼ	–	ん	ば	ぶ	べ	ば
M	ま	み	む	め	も	–	ん	ま	む	め	ま
R	ら	り	る	れ	ろ	っ	–	ら	る	れ	ら
W	わ	い	う	え	お	っ	–	わ	う	え	わ
SX	し	し	する	すれ	しよ	し	–	さ	する	–	さ
ZX	じ	じ	じる	じれ	じよ	じ	–	じら	じる	じれ	じさ
			ずる	ずれ			–		ずる		
KX	こ	き	くる	くれ	こよ	き	–	こら	くる	これ	こさ
KKX	来	来	来る	来れ	来よ	来	–	来ら	来る	来れ	来さ
RX	ら	り い	る	れ	ろ	っ	–	ら	る	れ	ら
IKU	か	き	く	け	こ	っ	–	か	く	け	か

("*" は語尾の表記文字がないことを表し，"–" はその活用形自体がないことを表す)

表 19　活用型と活用形

例えば「itarou」では，隣接型がT型で表記文字が「たろう」の形態素と，隣接型がD型で表記文字が「だろう」の形態素の二つに分ける．

これらの隣接規則は「動詞語幹の活用形名と，接尾辞の隣接型名が一致するものが隣接可能である」ということになる．例えば「書か」は動詞語幹「書」の活用形Aの形態であるから，隣接型がAの動形派生接尾辞「な」と隣接可能である．

4.2　活用形に対する追加

上記のような修正を語幹に対して行う場合，連体形，終止形，接続形の動詞接尾辞「ru」，連用形の動詞接尾辞「i」，さらに可能の動動派生接尾辞「re」は，動詞語幹の活用形として先頭の文字が吸収されてしまうと形態素としての表記文字が残らないという問題が起こる．また，

活用形 / 活用型	X	連用形	連体形	終止形	接続形	命令形
A	*	る	る	る	る	ろ，よ
K	*	き	く	く	く	け
G	*	ぎ	ぐ	ぐ	ぐ	げ
S	*	し	す	す	す	せ
T	*	ち	つ	つ	つ	て
N	*	に	ぬ	ぬ	ぬ	ね
B	*	び	ぶ	ぶ	ぶ	べ
M	*	み	む	む	む	め
R	*	り	る	る	る	れ
W	*	い	う	う	う	え
SX	–	し	する	する	する	しろ，せよ
ZX	じ	じ	じる	じる	じる	じろ
			ずる	ずる	ずる	ぜよ
KX	こ	き	くる	くる	くる	こい
KKX	来	来	来る	来る	来る	来い
RX	*	り	る	る	る	れ，い
IKU	*	き	く	く	く	け

表 20　活用形の追加

命令の動詞接尾辞の「ro」「e」には，動詞語幹の末尾が母音か子音かによって接続規則が異なるという問題がある．そこで動詞接尾辞の「ru」「i」「ro」「e」に関してはそれぞれ活用形としてしまう．そのため，表19に表20を加える．新たに加わったものは動詞語幹と動詞接尾辞が合成されたものであるので，属性も合成されたものになる．それを表21に示す．

可能の動動派生接尾辞「re」は，さらに後ろに動詞接尾辞などが来るため，活用形として加えられない．そこで，全ての活用型に対して語幹自体を活用形Xとして設定し，個々の子音との組み合わせによる「re」の変化を別々の形態素とした．そして，これらについて活用形Xに対する隣接規則をそれぞれ作ることで解決した．

このように，修正された接尾辞の扱いでは，一文字で構成される動詞接尾辞や派生接尾辞を新たに加えようとすると新たな活用形を作り出さなければならないが，一文字で構成されるという制約があるため，これ以上追加する必要が生ずる可能性は低い．

活用形	属性
連用形	[動, 連用, 無, 無]
連体形	[動, 連体, 無, 無]
終止形	[動, 終止, 無, 無]
接続形	[動, 無, 無, 接続]
命令形	[動, 終止, 無, 無]

表 21　合成された属性

4.3　形容詞に関する修正

前述したように形容詞の語幹は，形動派生接尾辞「ござr」が接尾する場合には連声する．しかし，これは非常に限られた現象で，しかもこれを本研究で使用した形態素解析プログラムの形態素文法に反映させると非常に煩雑になるので，実際にはこれを正確に実装はせずに「うござr」「ゅうござr」という形動派生接尾辞を辞書に登録した．こうすると「高ゅうござる」などを過剰に受理してしまい，また「たこうござる」のような平仮名表記の場合には解析ができない．過剰な受理に関しては，解析に対してなんらかの悪影響を及ぼさない限り許容する．実際，今までのところ，解析に関してはこのための悪影響は確認されていない．また，平仮名表記の場合は解析が不可能であるが，そのような事例は皆無に近いと考え，対処しないことにした．

5　問題点の検討

この節では以下に関する問題点について検討する．

- 複数の品詞に属する形態素
- 動名詞
- 連用詞に係る連用詞
- 複合名詞
- 複数解に対する優先度付け

複数の品詞に属する形態素の内，幾つかは形態素レベルの情報では識別できない．そのような形態素が現れる文は本来複数の解釈が存在し，これを完全に一つの解釈に決めるためには文脈を参照する必要がある．本研究での形態素解析システムでは，このような識別は形態素解析システムの範囲を超えるものと見なしている．以下でそのような形態素のついて述べるが，他の形態素解析システムとの性能の比較を容易にするために，新聞記事１万文中[9]の出現頻度についても述べ，正しい品詞を選ぶ確率が高くなるような規則を付す．

5.1　「名詞＋と」

9 形態素数約 20 万，文節数約 8 万 5 千

	格接尾辞連体形	格接尾辞連用形	引用接尾辞連用形	合計
名詞＋と	465(37.8%)	211(17.2%)	553(45.0%)	1229(100%)
名詞＋と＋名詞	436(35.4%)	102(8.3%)	8(0.7%)	546(44.4%)
名詞＋と＋動詞	0(0%)	93(7.6%)	519(42.2%)	612(49.8%)
名詞＋と＋引用性動詞	0(0%)	0(0%)	511(41.6%)	511(41.6%)
名詞＋と＋読点	20(1.6%)	4(0.3%)	4(0.4%)	28(2.2%)

表 22 「名詞＋と」の用法の分布

	名名派生接尾辞	引用接尾辞	格接尾辞	合計
名詞＋とも	46(63.9%)	18(25.0%)	8(11.1%)	72(100%)
名詞＋とも＋引用性動詞	0(0.0%)	18(25.0%)	0(0.0%)	18(25.0%)
文頭＋名詞＋とも	18(25.0%)	1(1.4%)	0(0.0%)	19(26.4%)
読点＋名詞＋とも	18(25.0%)	0(0.0%)	3(4.2%)	21(29.2%)
連用形＋名詞＋とも	7(9.7%)	5(6.9%)	0(0.0%)	12(16.7%)
連体形＋名詞＋とも	3(4.2%)	12(16.7%)	5(6.9%)	20(27.8%)

表 23 「名詞＋とも」の用法の分布

「名詞＋と」には，格接尾辞連体形 (名詞を並列に並べる用法) と，格接尾辞連用形 (共同作業者を示す用法) と，引用接尾辞連用形の三つがある．この内，最初の用法は連体であり，その他の用法は連用である．また，引用接尾辞の場合は述語を形成する．形態素レベルではこれらの用法を識別できない．表 22に「名詞＋と」の用法の分布を示す．なお，引用性動詞とは「なる」「する」「いう」「みる」「みなす」「思う」などあらかじめ選ばれた動詞である．この表によると，次の規則により，1060/1229(86.2%) の場合で正しい品詞を得られる．

- 「名詞＋と＋名詞」の場合は格接尾辞連体形
- 「名詞＋と＋引用性動詞」の場合は引用接尾辞連用形
- 「名詞＋と＋引用性動詞以外の動詞」の場合は格接尾辞連用形
- 「名詞＋と＋読点」の場合は格接尾辞連体形

5.2 「名詞＋との」

「名詞＋との」にも，格接尾辞連体形 (共同作業者を示す用法) と引用接尾辞連体形があり，前者は述語を形成しないが，後者は述語を形成する．この場合も，名詞に「との」が接尾しているものは，識別できない．「名詞＋との」は評価に用いた文中では 130 箇所に現れ，その内 123 箇所 (94.6%) が格接尾辞連体形であった．引用接尾辞連体形の場合は 7 箇所で，その係先は「認識」「見方」「情報」「主張」「理由」「考え」であり，逆にこれらの名詞に係る場合で格接尾辞であるものはなかった．

5.3 「名詞＋とも」

	格接尾辞連用形	名詞接尾辞連用形	合計
名詞＋で	2124(64.5%)	1171(35.3%)	3295(100%)
名詞＋で＋ある/ない	0(0.0%)	461(14.0%)	461(14.0%)
名詞＋で＋は/も＋ある/ない	0(0.0%)	134(4.1%)	134(4.1%)
名詞＋で＋は/も＋(ある/ない)以外	476(14.4%)	10(0.3%)	486(14.7%)
名詞＋で＋(ある/ない/読点)以外	1373(41.7%)	347(10.5%)	1720(52.2%)
連用形＋名詞＋で＋読点	256(7.8%)	129(3.9%)	385(11.7%)
連用形以外＋名詞＋で＋読点	19(0.6%)	90(2.7%)	109(3.3%)

表 24 「名詞＋で」の用法の分布

「名詞＋とも」は名名派生接尾辞，引用接尾辞＋連用接尾辞，格接尾辞＋連用接尾辞の三つの可能性がある．評価文中では名名派生接尾辞が 46 箇所，引用接尾辞＋連用接尾辞が 18 箇所，格接尾辞＋連用接尾辞が 8 箇所であった．表 23に「名詞＋とも」の用法の分布を示す．以下の規則により，66/72(91.7%) の場合で正しい品詞が得られる．

- 「名詞＋とも＋引用性動詞」の場合は，引用接尾辞．
- 「連体形＋名詞＋とも＋引用性動詞以外」の場合は，格接尾辞．
- 上記以外は，名名派生接尾辞．

5.4 「名詞＋で」

「で」には，格接尾辞連用形 (場所や道具を示す用法) と名詞接尾辞連用形がある．前者は述語を形成せず，後者は述語を形成する．これらは形態素レベルの情報では識別できない．「名詞＋で」の用法の分布を表 24に示す．これによると以下の規則で 2790/3295(84.7%) の場合で正しい品詞が得られる．

- 「名詞＋で＋ある/ない」の場合は，名詞接尾辞連用形
- 「名詞＋で＋は/も＋ある/ない」の場合は，名詞接尾辞連用形
- 「名詞＋で＋は/も＋(ある/ない)以外」の場合は，格接尾辞連用形
- 「名詞＋で＋(ある/ない/読点)以外」の場合は，格接尾辞連用形
- 「連用形＋名詞＋で＋読点」の場合は，名詞接尾辞連用形
- 「連用形以外＋名詞＋で＋読点」の場合は，格接尾辞連用形

5.5 「名詞＋か」

「か」には，格接尾辞連体形，名詞接尾辞連用形，接名派生接尾辞，接続接尾辞終止形がある．この内，「名詞＋か」では格接尾辞連体形と名詞接尾辞連用形が形態素レベルの情報では識別できない．ただし，本研究では例えば「太郎か次郎か分からない.」という文の場合，両方の「か」は名詞接尾辞連用形であると考えている．「名詞＋か」は評価の文中の 84 箇所に現れ，その内，10 箇所 (11.9%) が格接尾辞連体形，74 箇所 (88.1%) が名詞接尾辞連用形であった．ま

	補助名詞＋格接尾辞	接続接尾辞連用形	合計
のに	25(53.2%)	22(46.8%)	47(100%)
のに＋読点	1(2.1%)	14(29.8%)	15(31.9%)
のに＋名詞	15(31.9%)	5(10.6%)	20(42.6%)
のに＋述語	9(21.3%)	1(2.1%)	10(21.3%)

表 25　「述語＋のに」の用法の分布

た，評価文中では以下の規則で全ての場合を正しく識別できた．

- 「名詞＋か＋名詞」の場合は，格接尾辞連体形
- 「名詞＋か＋名詞以外」の場合は，名詞接尾辞連用形

5.6　「述語＋ので」

「述語＋ので」については，「の (補助名詞)＋で (格接尾辞)」，「ので (接続接尾辞連用形)」の二通りの解釈がある．例えば，「大きいので壊した.」という文では，「大きい物で壊した」のか「大きいから壊した」のかの区別ができない．しかし，前者の解釈は口語的なので，評価に用いた新聞記事では 1 万文の中に現れた 50 箇所全てが後者の用法であった．

5.7　「述語＋のに」

「述語＋のに」については，「の (補助名詞)＋に (格接尾辞)」，「のに (接続接尾辞連用形)」の二通りの解釈がある．例えば，「高いのに乗った.」では，「高いにも関わらず乗った」のか「高いものに乗った」のか識別できない．「述語＋のに」の用法の分布を表 25に示す．これによると，以下の規則で 38/47(80.9%) の場合に正しい品詞を得られる．

- 「のに＋読点」の場合は，接続接尾辞連用形
- 「のに＋名詞」の場合は，補助名詞＋格接尾辞
- 「のに＋述語」の場合は，接続接尾辞連用形

5.8　「そう」

「そう」には，動名派生接尾辞，形名派生接尾辞，名名派生接尾辞，接名派生接尾辞があり，ほとんどの場合はこれらは形態素レベルの情報で識別できる．しかし，「動詞語幹 (活用型 A) ＋そう」の場合には識別できない二つの解釈がある．例えば「食べたそうだ.」という文では「食べ (動詞語幹) た (動形派生接尾辞 [欲求]) そう (形名派生接尾辞 [様態]) だ (名詞接尾辞終止形)」と「食べ (動詞語幹) た (動詞接尾辞接続形 [完了]) そう (接名派生接尾辞 [伝聞]) だ (名詞接尾辞終止形)」を形態素レベルで識別できない．評価に用いた文中ではこのような「そう」は 11 箇所に現れ，その全てが後者の用法であった．これは評価に用いた文が新聞記事であるためと考えられる．

	動名詞	動詞	合計
動詞語幹＋ｉ＋名詞	79(60.5%)	50(39.5%)	129(100%)
読点＋動詞語幹＋ｉ＋名詞	13(10.0%)	0(0.0%)	13(10.0%)
名詞＋動詞語幹＋ｉ＋名詞	25(19.4%)	0(0.0%)	25(19.4%)
文頭＋動詞語幹＋ｉ＋名詞	6(4.7%)	0(0.0%)	6(4.7%)
連体形＋動詞語幹＋ｉ＋名詞	26(20.2%)	0(0.0%)	26(20.2%)
連用形＋動詞語幹＋ｉ＋名詞	9(7.0%)	50(39.5%)	59(45.7%)
は＋動詞語幹＋ｉ＋名詞	7(5.4%)	0(0.0%)	7(5.4%)

表 26 「動詞語幹＋ｉ＋名詞」の用法の分布

5.9 「動詞語幹＋ｉ＋に」

「動詞語幹＋ｉ＋に」には二つの解釈がある．例えば「話しに花を添える.」「話しに行く.」では「話し」は前者では「動名詞＋格接尾辞 “に”」であり，後者では「動詞語幹＋動詞接尾辞 “ini”」である．これらは形態素レベルの情報では区別できない．これは評価文中では 168 箇所で現れた．その内 150 箇所が動名詞であり，18 箇所が動詞であった．動詞の 18 箇所の内,「動詞＋に＋行く」が 5 箇所であり,「動詞＋に＋来る」が 8 箇所，その他,「入る」「通う」「向かう」「寄る」が直後に来るものがそれぞれ 1 箇所ずつあった．逆にこれらの動詞が直後に来る場合で動名詞であったものはなかった．従って，以下の規則で 167/168(99.4%) が正しく識別できる．

- 「来る」「行く」などの特定の動詞が直後に来る場合は「動詞語幹＋動詞接尾辞 “ini”」.
- 上記以外の場合は「動名詞＋格接尾辞 “に”」.

5.10 「動詞語幹＋ｉ」

「動詞語幹＋ｉ」には，動詞の連用形である場合と，動名詞である場合がある.「動詞語幹＋ｉ＋名詞接尾辞」「動詞語幹＋ｉ＋格接尾辞」「動詞語幹＋ｉ＋連用接尾辞」の場合は動名詞であると識別することができる．また「動詞語幹＋ｉ＋読点」は動詞の連用形と識別できる．

文節に区切る際に最も問題になるのは「動詞語幹＋ｉ＋名詞」の場合である.「動詞語幹＋ｉ」を動詞の連用形と見なす場合にはそこで文節が区切れるが,「動詞語幹＋ｉ」を動名詞と見なす場合には複合名詞になるので文節が区切れない．このような「動詞語幹＋ｉ＋名詞」のパターンは評価文中の 129 箇所に現れ，動名詞であったのが 79 箇所であり，動詞であったのが 50 箇所であった．これらは形態素レベルの情報では区別できない．しかし，評価文中では表 26のような用法の分布があった．従って，下記の規則で 127/129(98.4%) の場合で正しく識別できる．

- 「連用接尾辞 “は” ＋動詞語幹＋ｉ＋名詞」の場合は，動名詞.
- 「連用接尾辞 “は” 以外の連用形＋動詞語幹＋ｉ＋名詞」の場合は，動詞.
- 上記以外は動名詞.

5.11　「いく」と「いう」

「いった」「いって」などは「言った」「言って」なのか「行った」「行って」なのか分からない．評価の文中には77箇所でこのような表現が現れたが，動詞の連用形の直後に来るものは全て「行く」であり，それ以外はすべて「言う」であった．これは，評価に用いた文が校正済みの新聞記事であるためと考えられる．

5.12　「ある」

「ある」には，連体詞と動詞の可能性がある．これは評価文中に260箇所に現れ，24箇所が連体詞，236箇所が動詞であった．この内,「読点＋ある」は8箇所で，全て連体詞であった．その他の連体詞の「ある」は16箇所全てが「名詞＋の＋ある」の形で現れたが，動詞の「ある」が「名詞＋の＋ある」の形で現れたのが29箇所であった．従って，以下の規則で，244/260(93.8%)が正しく認識される．

- 「連用形＋ある」の場合は，動詞．
- 「名詞＋の＋ある」の場合は，動詞．
- 上記以外の「連体形＋ある」の場合は，連体詞．
- 「読点＋ある」の場合は，連体詞．

5.13　連用詞に係る連用詞

連用詞の中には他の連用詞を修飾していると考えられるものがある．例えば「非常にゆっくり歩いた.」という文で,「非常に」は「ゆっくり」の様態を表していると考えられる．一つの解決法は連用詞は他の連用詞に係ることができるとしてしまうことであるが，全ての連用詞が他の連用詞に係るわけではないので，連用詞に係ることができる連用詞として別の品詞を設定する必要が出てくる．別の解決法は，先ほどの例で言えば,「非常に」が「ゆっくり」ではなく「歩いた」に係ると見なすことにしてしまう方法である．その場合は,「非常に」と「ゆっくり」の関係を「歩いた」を仲介して算出する仕組みを別に用意しなければならない．しかし，この利点は,「非常に私はゆっくり歩いた.」という文でも係り受けの非交差の原則が守られていると見なせる点である[10]．また連用詞の変種を作る必要もないので，本研究では後者の解決法を取っている．

10 同様な現象は「は」にも見られ，例えば「この料理は私は彼女が作ったと思う.」という文で「料理は」が「作った」に係るとすると非交差の原則が破られるが，これも「料理は」は「思う」に係ると見なして，別の仕組みによって,「料理」と「作った」の関係を算出すると考えれば，非交差の原則が守られていると見なせる．

5.14 複合名詞

本研究における形態素解析システムは，その目的から，複合名詞をさらに細かく区切ることを重要視していない．つまり，文節の区切りの精度を測定する場合に「名詞＋名詞[11]」の並びを複合した結果が名詞として正解であればよしとしている．従って，複合した状態では正解であっても，それをさらに細かく分解した状態では間違っている場合がある．評価文中では11845箇所に複合名詞の分割が現れた[12]．この内，492箇所(4.2%)が誤って分割されていた．誤りの内，221箇所(1.9%)が固有名詞に起因するものであった．

5.15 複数解に対する優先度付け

本稿では述べていないが，実際の形態素解析処理における重要な要素に複数解に対する優先度付けの問題がある．例えば，「太郎が帰ってきたとき，犬が吠えた.」という文には，本稿で示した形態素文法だけでは，「とき」の部分に曖昧性が生じる．一つの解釈は明らかなように「とき(時)」という名詞である．今一つの解釈は，「と(引用)」「き(“来る”の連用形)」である．ここでは明らかに前者の解釈を取らなければならない．その他にも，単語の平仮名表記を含めれば多くの曖昧性がある．そこで，実際の形態素文法の定義では，品詞や品詞の隣接規則に重み付けをし，優先度の計算を行っている．しかし，この重み付けはまったくアドホックなものであり，実際，多くの例文を処理させた結果を分析して，優先度の計算がうまく人間の解釈と適合するように調整する事によって作成した．実際のシステムへの適用にあたってはこの部分が最も時間がかかった部分であり，さらなる精度向上に対する障害の一つである．

6 性能評価

本稿で提案した形態素文法を形態素解析プログラムJUMAN(松本他 1994)[13]に適用し，形態素解析の精度を測定した．本来のJUMANは接続コストによって枝狩りした解に対して後方最長一致の解を出力するもの[14]であるが，本研究ではこれを接続コストが最小になる解(久光・新田 1990)を出力するように改造して使用した．利用した辞書は異なり語数３５万程度であり，これらの内，動詞語幹，形容詞語幹，名詞，連用詞，連体詞，名名派生接尾辞，数名派生接尾辞については日本電子化辞書研究所の日本語辞書の他，いくつかの仮名漢字変換プログラム用辞書や機械可読な人間用の辞書から抽出したものを用いた．また，漢字表記の語については，その平仮名表記も辞書に登録し，全体で５０万語程度となっている．ただし，単語の中で漢字の一部だけを平仮名に変えたものは辞書に登録していない．

11 「連用名詞＋名詞」は複合名詞にならない．
12 この中には辞書に一つの名詞として登録されている複合名詞は含まれない．そのような複合名詞は1302箇所に現れた．
13 JUMANは品詞や形態素文法を再定義できる公開された形態素解析システムである．
14 オプションによってただ一つの解を出力するように指定した場合

文数	10000
文節数	84841
形態素数	207547
文節区切り位置の誤り数	445
分割誤り複合名詞数	221

区切り誤り

	頻度	誤り数
名詞＋と	1229	169
名詞＋との	130	0
名詞＋とも	72	6
名詞＋で	3295	505
名詞＋か	84	0
述語＋ので	50	0
述語＋のに	47	9
動詞語幹 (A) ＋そう	11	0
動詞語幹＋ｉ＋に	168	1
動詞語幹＋ｉ＋名詞	129	2
ある	260	16

品詞付け誤り

表 27　誤り数

評価には日本電子化辞書研究所から提供されたコーパス[15]の内，1 万文を使用した．これらの文に対して形態素解析システムに**ただ一つ** の解を出力させ，これとコーパスに付けられている人手による解析結果とを比較した．ただし，日本電子化辞書研究所における品詞の分類と本論文での品詞の分類が異なっているため，文節単位にまで形態素をまとめたものを比較した．結果を表 27に示す．但し，「区切り誤り」は 5節で与えた規則によって品詞が間違う場合の数である．

文節の区切り位置を誤っていたのは 445 箇所であった．区切り位置を誤ると，その前後の文節が共に誤りとなるので，誤りの文節が含まれる率は，全形態素数に対して，

$$445 \times 2 \div 207547 \times 100 = 0.43\%$$

である．これは全文節数に対しては 1.05%である．また，1 文中に複数の区切り誤りがあったものはなかったので，文節区切りが失敗した文は全文数に対して 4.45%である．

文献 (丸山・荻野 1994) では，分割誤りとして複合名詞の分割誤りを含めて，形態素数に対して分割誤り率は 1.25%と報告されている．本稿のシステムを同様に評価すると，

$$(445 + 221) \times 2 \div 207547 \times 100 = 0.64\%$$

の分割誤り率である．さらに文献 (丸山・荻野 1994) では品詞誤りを含めた全体的な誤り率を 2.36%と報告している．文献 (丸山・荻野 1994) では本稿で与えた形態素文法よりも細かい品詞分類を行っているので，同様に比較できないが，表 27に挙げたものを形態素に対する品詞付けの誤りとすると，

$$((445 + 221) \times 2 + 169 + 6 + 505 + 9 + 1 + 2 + 16) \div 207547 \times 100 = 0.98\%$$

である．

15 このコーパスには主に朝日新聞社の記事から収集した文が集められている．

文節の区切りに関する誤りは，8箇所が複数解の優先度付けの誤りによるものであり，残りの437箇所は対応する形態素を辞書に登録することによって解決できるものであった．従って，辞書を整備することで文節区切りの性能はさらに向上させることができると期待できる．辞書の整備に関して，語の中の一部の漢字が平仮名表記されるものについては，自動的に漢字の一部を平仮名に置き換えたものを登録することが可能である．しかし，その場合，登録語数がほぼ4倍になる．実際にはこれらの内ほとんどのものは用いられない上，解析速度にも悪影響を与えるので，コーパスの分析結果などから必要な表記のみを登録するのが望ましい．

複合名詞の区切り誤りについては本形態素文法では対処できない．しかし，複合名詞としてまとまった形で認識する精度は高い．複数の品詞に属する形態素に関しては，新聞記事に対しては有効性の高い識別規則を与えたが，これらはあくまで確率的なものであり，根本的な解決にはならない．

動詞の語尾変化は全て正しく解析され，派生文法における動詞の取り扱い方法の優秀さが実証された．口語的な表現に対しても，JUNETの生活関連のニュースグループの記事の内，明らかな間違いを除いた500文を解析させたところ，動詞語の語尾変化に対しては全て正しく解析されていた．

7　まとめ

本稿では形態素解析に的を絞った日本語形態素文法を提案した．この形態素文法における動詞語尾の扱いは，派生文法を拡充整備し，日本語の文字単位で扱えるように修正したものである．その結果，実存する形態素解析プログラムJUMANに適用できるようになり，実際に適用して実用的な解析性能を得ることができた．辞書を整備することでさらなる精度の向上が期待できる．しかし，形態素の隣接規則間の優先度を決める重みの決定は，手作業による微妙な調整によるものであり，何らかの自動的な学習の仕組みが必要である．

謝辞

形態素解析プログラムJUMANを提供して下さった奈良先端科学技術大学院大学の松本裕治先生，および辞書を提供して下さった日本電子化辞書研究所の方々に感謝します．

参考文献

Bloch, B. (1946). "Studies in Colloquial Japanese, Part I, Inflection." *Journal of the American Oriental Society*, **66**.

Hisamitu, T. and Nitta, Y. (1994). "An Efficient Treatment of Japanese Verb Inflection for Morphological Analysis." In *Coling 94*, Vol. I.

久光徹, 新田義彦 (1990). “接続コスト最小法による日本語形態素解析の提案と計算量の評価について.” 言語理解とコミニュケーション 90-8, 電子情報通信学会.

久光徹, 新田義彦 (1994a). “ゆう度付き形態素解析用の汎用アルゴリズムとそれを利用したゆう度基準の比較..” 電子情報通信学会論文誌 D-II, **J77** (5).

久光徹, 新田義彦 (1994b). “日本語形態素解析における効率的な動詞活用処理.” 自然言語処理研究会 103-1, 情報処理学会.

木谷強 (1992). “固有名詞の特定機能を有する形態素解析処理.” 自然言語処理研究会 90-10, 情報処理学会.

清瀬 義三郎則府 (1989). 日本語文法新論 –派生文法序説–. 桜楓社.

丸山宏, 荻野紫穂 (1994). “正規文法に基づく日本語形態素解析.” 情報処理学会論文誌, **35** (7).

益岡隆志, 田窪行則 (1992). 基礎日本語文法 –改訂版–. くろしお出版.

松本裕治, 黒橋禎夫, 宇津呂武仁, 妙木裕, 長尾真 (1994). “日本語形態素解析システム JUMAN 使用説明書 version 2.0.” テクニカル・レポート, 奈良先端科学技術大学院大学.

三浦つとむ (1975). 日本語の文法. 勁草書房.

宮崎正弘, 高橋大和 (1992). “三浦文法に基づく日本語形態素処理用文法の構築.” 自然言語処理研究会 90-1, 情報処理学会.

Nagata, M. (1994). “A Stochastic Japanese Morphological Analyzer Using a Forward-DP Backward-A* N-Best Search Algorithm.” In *Coling 94*, Vol. I.

中村順一, 吉田将, 今永一弘 (1991). “接続コスト最小法による日本語形態素解析の評価実験.” 言語理解とコミニュケーション 91-1, 電子情報通信学会.

西野博二, 鷲北賢, 石井直子 (1992). “派生文法による日本語構文解析.” 自然言語処理研究会 87-6, 情報処理学会.

寺村秀夫 (1984). 日本語のシンタクスと意味 II. くろしお出版.

時枝誠記 (1950). 日本語文法 口語篇. 岩波書店.

吉村賢治, 日高達, 吉田将 (1983). “文節数最小法を用いたべた書き日本語の形態素解析.” 情報処理学会論文誌, **24** (1).

略歴

渕 武志: 1965 年生. 1988 年東京大学理学部情報科学科卒業. 1991 年慶応大学大学院修士課程終了. 1995 年東京大学大学院博士課程修了. 理学博士. 同年, ＮＴＴに入社, 現在に至る. 自然言語処理, 知識情報処理の研究に従事.

米澤明憲: 1947 年生. 1977 年 Ph.D. in Computer Science (MIT). 1989 年より東京大学理学部情報科学科教授. 超並列ソフトウエアアーキテクチャ, ソフトウエア基礎論, 人工知能基礎論などに興味を持つ. 著書「算法表現論」, 「モデルと表現」(岩波書店), 編著書「ABCL: An Object-Oriented Concurrent

System」,「Research Directions in Concurrent Object-Oriented Computing」(MIT Press) 等. 現在 IEEE Parallel and Distributed Technology 編集委員. 1992 年よりドイツ国立情報処理研究所 (GMD) 科学顧問.

(1994 年 10 月 21 日 受付)
(1995 年 1 月 10 日 再受付)
(1995 年 3 月 6 日 再々受付)
(1995 年 3 月 23 日 採録)